

- 梯度下降
  - 一、什么是梯度
  - 二、梯度下降的例子：下山
  - 三、梯度下降在机器学习的用途

# 梯度下降

## 一、什么是梯度

多元函数的各参数求偏导数，然后把所求得各参数的偏导数以向量的形式写出来，就是梯度。

示例：两个自变量的函数  $z = f(x, y)$  对应着机器学习数据集中的两个特征，如果分别对  $x$  和  $y$  求偏导，那么求得的梯度向量就是  $\begin{bmatrix} \frac{\partial z}{\partial x} \\ \frac{\partial z}{\partial y} \end{bmatrix}$ ，在数学上表示为  $\nabla f(x, y)$ 。

梯度向量的意义：几何意义就是函数变化的方向，而且是变化最快的方向。对于函数  $z = f(x, y)$ ，在点  $(x_0, y_0)$  处，梯度向量的方向也就是  $z$  值增加最快的方向。也就是沿着梯度向量的方向  $z$  能找到函数的最大值。

反过来说，如果沿着梯度向量的反方向，梯度减少最快，能找到函数的最小值。

如果某一个点的梯度向量的值为0，那么也就是来到了导数为0的函数的最低点或者局部最低点。

## 二、梯度下降的例子：下山

由下山的样例，可以理解，函数的凹凸性对梯度的影响。在非凸函数中有可能还没有走到山脚，而是到了某个山谷就停住了，也就是对应非凸函数梯度下降不一定总能找到全局最优解，有可能得到的知识一个局部最优解。

但是，如果是凸函数，那么梯度下降法理论上就能得到全局最优解。

## 三、梯度下降在机器学习的用途

关注如下一些知识：

- 机器学习的本质是找到最优函数。
- 如何衡量最优？尽量减少预测值和真实值之间的误差，又叫损失值。
- 可以建立误差和模型参数之间的函数（最好是凸函数）
- 梯度下降能够引导我们走到凸函数的全局最低点，也就是误差最小的参数。